

# Probabilistic Computation in Human Perception under Variability in Encoding Precision

Shaiyan Keshvari<sup>‡</sup>, Ronald van den Berg, Wei Ji Ma<sup>\*</sup>

Department of Neuroscience, Baylor College of Medicine, Houston, Texas, United States of America

## Abstract

A key function of the brain is to interpret noisy sensory information. To do so optimally, observers must, in many tasks, take into account knowledge of the precision with which stimuli are encoded. In an orientation change detection task, we find that encoding precision does not only depend on an experimentally controlled reliability parameter (shape), but also exhibits additional variability. In spite of variability in precision, human subjects seem to take into account precision near-optimally on a trial-to-trial and item-to-item basis. Our results offer a new conceptualization of the encoding of sensory information and highlight the brain's remarkable ability to incorporate knowledge of uncertainty during complex perceptual decision-making.

**Citation:** Keshvari S, van den Berg R, Ma WJ (2012) Probabilistic Computation in Human Perception under Variability in Encoding Precision. PLoS ONE 7(6): e40216. doi:10.1371/journal.pone.0040216

**Editor:** Marc O. Ernst, Bielefeld University, Germany

**Received:** November 2, 2011; **Accepted:** June 6, 2012; **Published:** June 29, 2012

**Copyright:** © 2012 Keshvari et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** WJM is supported by award number R01EY020958 from the National Eye Institute (<http://www.nei.nih.gov/>). RvdB is supported by the Netherlands Organisation for Scientific Research (<http://www.nwo.nl>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [wjma@bcm.edu](mailto:wjma@bcm.edu)

‡ Current address: Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

## Introduction

The sensory information used by the brain to infer the state of the world is noisy: when the same stimulus is presented repeatedly, the neural activity it elicits varies considerably from trial to trial [1,2,3,4]. As a consequence, an observer's measurement of a task-relevant stimulus feature varies as well. The quality of the sensory information can be numerically expressed as precision. For instance, when the measurement follows a Gaussian distribution, precision could be defined as the inverse of the variance of this Gaussian.

Models of perception routinely assume that the precision with which a task-relevant stimulus feature is encoded is constant as long as the stimulus is held constant [5]. It is questionable, however, whether this assumption is justified, considering that factors such as fluctuations in alertness [6], configural effects [7,8], and covert shifts of attention [9,10] could make precision variable. If all factors were known and quantifiable, encoding precision could be specified exactly for each stimulus on each trial. However, as long as we are not able to model each possible contributing factor, it may be best to model precision as a random variable [11]. For example, the inverse variance of a Gaussian noise distribution could be drawn from a gamma distribution.

If encoding precision is a random variable, then the measurement of a task-relevant stimulus feature follows a doubly stochastic process. This idea translates to the level of neural coding, where a population pattern of activity could be Poisson-like with a mean amplitude (gain) that itself follows some other distribution. Recent physiological studies have reported evidence for doubly stochastic processes in cortex [12,13,14,15].

In the optimal-observer models of many tasks, precision does not only appear as part of the encoding model (a description of how measurements are generated), but also in the observer's decision rule (a description of how measurements are transformed into a decision). In other words, in some tasks, in order to be optimal, an observer must take into account precision even if precision varies unpredictably across stimuli and trials. To distinguish this type of computation from computation in which the observer can be optimal using only a point estimate of each stimulus feature, we use the term "probabilistic computation" [16]. At the neural level, probabilistic computation suggests that populations of neurons encode and compute with probability distributions over stimulus features [16,17,18], instead of only point estimates.

Psychophysical evidence for probabilistic computation has been found in cue combination tasks [19,20,21] as well as more complex categorization tasks [22,23]. In these experiments, the encoding precision of the task-relevant feature was manipulated by varying a reliability parameter, for example the size of a blurred disc if its location is task-relevant, or contrast of a bar if its orientation is task-relevant. Since we propose here that factors other than this reliability parameter also contribute to variability in precision, the question arises whether observers optimally take into account this additional variability.

Here we use a visual change detection task [24,25,26] to study whether precision is variable for a given value of the reliability parameter and whether observers take any variability in precision (whether or not due to the reliability parameter) into account optimally. Observers reported whether a change in the orientation of a stimulus occurred between two displays that each contained four stimuli (items). The reliability of the orientation information

was controlled by shape and was randomly chosen for each stimulus. We pitted an optimal-observer model in which precision is completely determined by shape (“equal precision”) against one in which there is additional variability (“variable precision”). Both models assume that precision is known and optimally taken into account by the observer on an item-by-item and trial-by-trial basis. We compare these two models to several suboptimal models, where suboptimality can be caused by two factors. First, the observer might make a wrong assumption about precision. For example, if precision varies across stimuli at different locations, the observer might assume a single value of precision for all stimuli instead of using the individual values. Second, the observer might use a suboptimal decision rule instead of the optimal rule to integrate information from different locations. Considering all combinations of model elements – equal or variable precision, various observer assumptions about precision, and two possible integration rules – we arrive at a total of 14 models. We find that the empirical data for each individual subject are best described by the model in which precision is variable, the observer knows precision on an item-by-item and trial-by-trial basis, and uses the optimal integration rule.

## Results

### Experiment

Subjects were presented with two consecutive displays, each presented for 100 ms and separated by a 1-second blank screen. Each display contained a set of four randomly oriented ellipses that were identical between both displays except that with 50% probability, exactly one ellipse changed orientation between the first and the second screen (Fig. 1A). The magnitude of a change, if present, was drawn from a uniform distribution. On each trial, we first randomly chose the number of high-reliability stimuli (0 to 4, with equal probability); then, we randomly chose which of the stimuli had high reliability. Reliability was controlled by shape: high-reliability ellipses were more elongated than low-reliability ones, but had the same area. Subjects indicated whether or not a change occurred.

As expected, subjects became better at detecting a change as the number of high-reliability stimuli, denoted  $N_H$ , increased (Fig. 1B). While we did not find a significant effect of  $N_H$  on the false-alarm

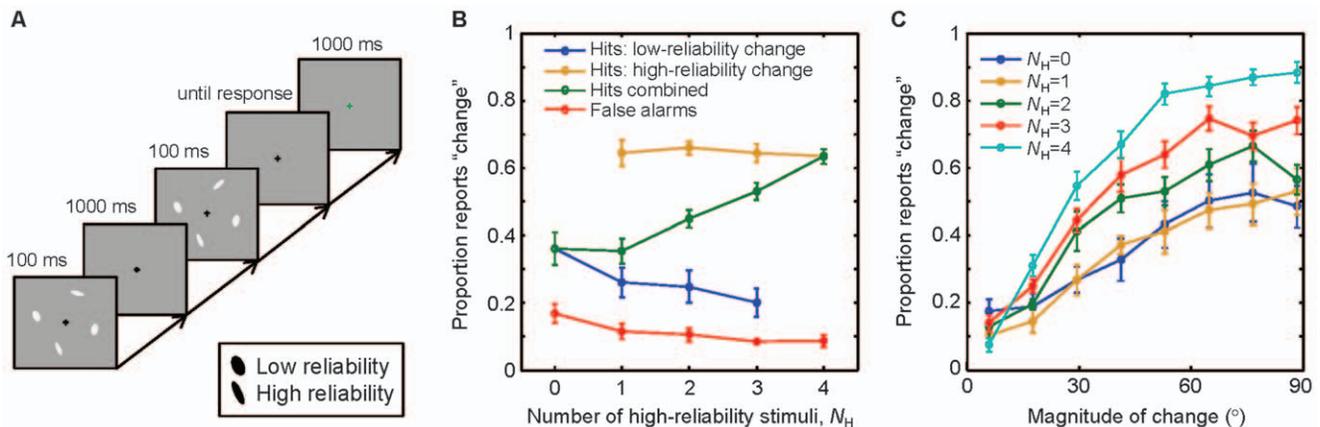
rate (one-way repeated-measures ANOVA,  $F(2.3,18.6)=2.9$ ,  $p=0.08$ ; degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity), the effect of  $N_H$  on the hit rate was significant ( $F(1.7,13.9)=25.1$ ,  $p<0.001$ ). This shows that our reliability manipulation was effective. Mean accuracy exceeded chance at every value of  $N_H$  ( $t(8)>5.5$ ,  $p<10^{-3}$ ).

When we separate hit trials by the reliability of the changing stimulus, we see a distinctive Z-shaped pattern (Fig. 1B). The hit rate conditioned on the change being in a low-reliability stimulus decreases monotonically with  $N_H$  ( $F(3,24)=9.7$ ,  $p<0.001$ ). We did not find an effect of  $N_H$  on the hit rate conditioned on the change being in a high-reliability stimulus ( $F(1.4,11.6)=0.20$ ,  $p=0.75$ ). It might be counterintuitive that the low-reliability hit rate decreases and the high-reliability hit rate is flat, yet the unconditioned hit rate increases. This effect is an instance of Simpson’s paradox [27]. The apparent contradiction is resolved by realizing that the relative contributions of the conditional rates change with  $N_H$ : the higher  $N_H$ , the larger the proportion of trials that fall in the high-reliability-change category. The Z-shaped pattern in our data confirms a prediction from an optimal model of a change discrimination task [28] (elaborated below).

Next, we binned change trials by magnitude of change (8 bins) (Fig. 1C). A two-way repeated-measures ANOVA reveals significant main effects of magnitude of change ( $F(7,56)=109.0$ ,  $p<0.001$ ) and of  $N_H$  ( $F(1.9,15.2)=24.4$ ,  $p<0.001$ ) on the proportion of “change” reports, and a significant interaction ( $F(28,224)=5.4$ ,  $p<0.001$ ). This indicates that larger changes are easier to detect.

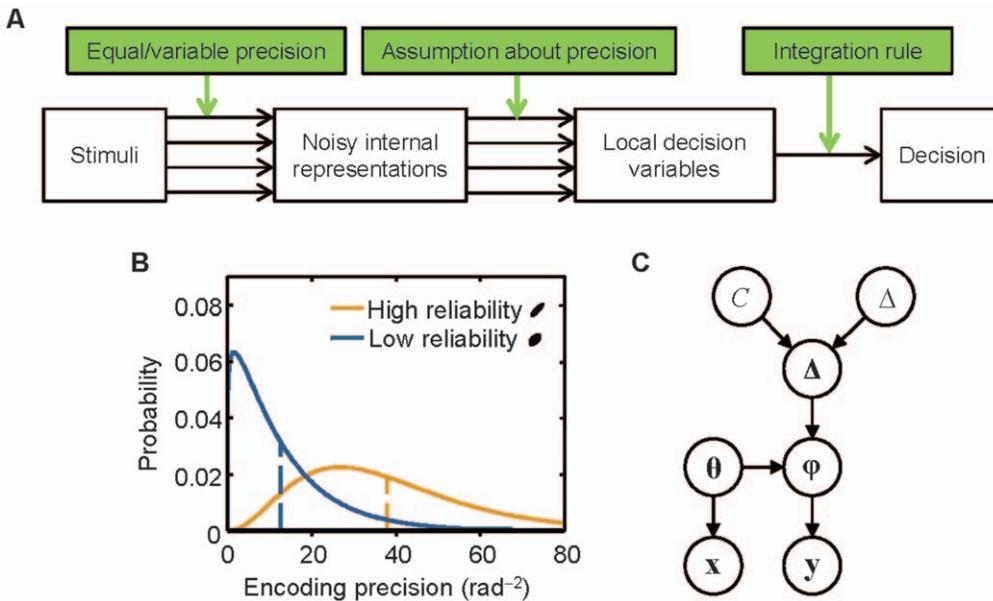
### Models

We model the observer’s decision process as consisting of an encoding stage and a decision stage (Fig. 2A). In the encoding stage, precision is either completely determined by stimulus reliability (“equal precision” or EP), or a random variable itself (“variable precision” or VP). Precision is technically defined as Fisher information (see Methods) and denoted  $\mathcal{J}$ . For a given value of precision,  $\mathcal{J}$ , the measurement  $x$  of an orientation  $\theta$  follows a probability distribution  $p(x|\theta;\mathcal{J})$ . For this distribution, we assume a circular Gaussian (Von Mises) distribution, characterized by a concentration parameter  $\kappa$  that corresponds one-to-one with precision (see Text S1 and Fig. S1). When precision is variable



**Figure 1. Change detection under varying reliability.** A, Schematic of the trial procedure. Stimulus reliability was controlled by ellipse elongation. Set size was always 4. B, Hit and false-alarm rates as a function of the number of high-reliability stimuli (long ellipses),  $N_H$ . Hit rates are split out by whether the changing ellipse had high or low reliability. The Z-shape formed by the yellow, green, and blue lines is an instance of Simpson’s paradox (see Results). C, Proportion of “change” reports in change trials as a function of the magnitude of change, for different values of  $N_H$ . Error bars represent  $\pm 1$  s.e.m.

doi:10.1371/journal.pone.0040216.g001



**Figure 2. A, Flow diagram of the decision process.** Models differ along three dimensions: whether precision is equal or variable, the observer’s assumption about precision, and the observer’s integration rule. **B**, Examples of probability density functions over encoding precision for a high-reliability and a low-reliability stimulus (long and short ellipse, respectively) in the variable-precision model. Dashed lines indicate the means. **C**, The generative model shows statistical dependencies between variables. **C**: change occurrence (0 or 1);  $\Delta$ : magnitude of change;  $\Delta$ : vector of change magnitudes at all locations;  $\theta$  and  $\phi$ : vectors of stimulus orientations in the first and second displays, respectively;  $\mathbf{x}$  and  $\mathbf{y}$ : vectors of measurements of the stimulus orientations. The spatial, temporal, and structural complexities of the task can be recognized in the vector nature of the orientation variables, the two “branches”, and the number of layers, respectively. doi:10.1371/journal.pone.0040216.g002

(VP), the measurement of a stimulus over many trials is described by a doubly stochastic process, formalized as the following integral:

$$p(x|\theta) = \int p(x|\theta; J)p(J)dJ, \tag{1}$$

where  $p(x|\theta; \mathcal{J})$  is again the Von Mises distribution and the variability in  $\mathcal{J}$  itself,  $p(\mathcal{J})$ , is modeled as a gamma distribution (Fig. 2B). The distribution in Eq. (1) is a mixture of an infinite number of Von Mises distributions, each with its own precision; it is a circular analog of the Student t-distribution.

In the decision stage, the Bayes-optimal observer computes on each trial the probability that a change occurred and responds “change” if this probability is greater than 0.5. This is equivalent to responding “change” when.

$$\frac{p_{\text{change}}}{1 - p_{\text{change}}} \frac{1}{N} \sum_{i=1}^N d_i > 1, \tag{2}$$

where  $p_{\text{change}}$  is the observer’s prior belief that a change occurred,  $N$  is the number of stimuli, and  $d_i$  is the local decision variable (i.e., the posterior probability ratio of change occurrence at the  $i^{\text{th}}$  location, denoted  $d_i$ ; see Text S1 for derivation).

$$d_i = \frac{I_0(\kappa_{x,i})I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(y_i - x_i)}\right)} \tag{3}$$

where  $x_i$  and  $y_i$  are the measurements of the  $i^{\text{th}}$  stimulus in the first and second displays, respectively,  $\kappa_{x,i}$  and  $\kappa_{y,i}$  are the corresponding concentration parameters of the noise, and  $I_0$  is the modified Bessel function of the first kind of order 0. Eq. (3) represents “weighting” by encoding precision (through  $\kappa_{x,i}$  and  $\kappa_{y,i}$ ) on a trial-by-trial and item-by-item basis, in a way analogous to but more complex than cue combination. It is crucial that the optimal observer knows precision,  $\mathcal{J}$ , and therefore  $\kappa$ , for each display and each item on each trial. Thus, even though Eq. (1) describes a doubly stochastic process over many trials, the optimal observer on a single trial knows the exact conditioned distribution  $p(x|\theta, \mathcal{J})$ .

In the decoding stage, the models we consider differ along two dimensions that can be understood in the context of Eqs. (3) and (2), respectively. The first dimension concerns the assumption that the observer makes about encoding precision:

1. no assumption: complete knowledge of an item’s precision on each trial, i.e. the optimal model;
2. the assumption that precision is completely determined by shape, ignoring any other variability (suboptimal);
3. the assumption that precision is equal to the average precision across the display (which will vary across trials), reflecting a “gist” representation of precision (suboptimal);
4. the assumption that precision is equal throughout the experiment, thus ignoring both variations in shape and other variability (suboptimal).

If encoding precision is equal (EP), assumptions 1 and 2 are equivalent, because there is no additional variability to ignore. Assumptions 2 to 4 are formalized as variants of Eq. (3) in which the trial-to-trial and item-to-item concentration parameters are replaced by values that are solely determined by stimulus

reliability, by the average value in the display, or by a single value throughout the experiment, respectively.

The second dimension along which the models differ is the integration rule that the observer applies to the local decision variables,  $d_i$ . Specifically, besides the optimal rule, Eq. (2), we consider the suboptimal “Max” rule, according to which the observer responds based on the largest local decision variable. The Max decision rule is  $\max d_i > k$ , with  $k$  a constant criterion. The Max rule has been used widely in signal detection theory models of visual search and is considered a reasonable description of human search behavior [29,30,31,32] (but see [22]). The Max model together with the assumption of single precision (Assumption 4) is equivalent to the (also suboptimal) maximum-absolute-differences model we introduced for change detection in earlier work [33] (see Text S1). In total, this produces  $(4+3)\cdot 2 = 14$  models, listed in Table 1. The number of free parameters ranges from 3 to 5.

### Model Comparison

We compared the models in two ways. First, we fitted each model’s parameters using maximum-likelihood estimation and computed  $R^2$  for the fits to the data in Fig. 1B-C (Fig. 3). The winning model was the one in which encoding precision is variable, observers optimally weight observations by their encoding precision, and they use the optimal rule for integrating information across locations (the VVO model from Table 1). This model had the highest goodness-of-fit for hit and false-alarm rates ( $R^2 = 0.97$ ), as well as for psychometric curves ( $R^2 = 0.89$ ). Maximum-likelihood estimates of model parameters are given in Table S1.

Second, to distinguish the models in a more powerful way, we performed Bayesian model comparison [34]. This method computes the average likelihood over all parameter combinations, thereby automatically correcting for the number of free parameters (see Online Methods). The VVO model is the clear winner for each of the 9 subjects individually. Bayesian model comparison revealed that the log likelihood of the VVO model exceeds that of the next best model (VVM, which uses the Max rule) by the decisive difference of  $15.4 \pm 17.3$  (mean and s.e.m.) log likelihood points (Fig. 4).

The VVO model exceeds the EEO model – the best equal-precision model – by  $36.3 \pm 6.3$  log likelihood points, suggesting variability in encoding precision. To confirm that this advantage is not due to unmodeled noise at the decision level (the last two steps in Fig. 2A), we tested two EEO model variants that included such noise. In the first variant (“local decision noise”), we added zero-mean Gaussian noise with standard deviation  $\sigma_{\text{local}}$  to the log of the local decision variable,  $d_i$ . In the second variant (“global decision noise”), we added the same type of noise (with standard deviation  $\sigma_{\text{global}}$ ) to the log of the left-hand side of Eq. (2). The best-fitting values were  $\sigma_{\text{local}} = 0.34 \pm 0.04$  and  $\sigma_{\text{global}} = 0.30 \pm 0.08$ . These values are small given that log decision variables generally ranged from about  $-4$  to  $20$ . Furthermore, we computed the model likelihoods of these two variants, and compared them to that of the winning model, VVO. The EEO models with local and global decision noise had log likelihoods of  $-37.1 \pm 7.0$  and  $-38.2 \pm 7.0$  relative to VVO, respectively. Moreover, the VVO model described the data better than both noisy models in all nine subjects individually. Thus, decision noise cannot account for the difference between the VVO and EEO models.

### Simpson’s Paradox

As Fig. 3A shows, the VVO model accounts for the characteristic Z-shape in the hit rates. The intuition behind the Z-shape in the context of the VVO model – and in fact any model

that weights observations by their encoding precision – is as follows. The unconditioned hit rate increases with the number of high-reliability stimuli,  $N_H$ , because more information is available in the measurements, and the observer utilizes this information. The hit rate conditioned on the changing item having low reliability decreases with increasing  $N_H$  because a higher value of  $N_H$  means that more non-changing items have high reliability. Since in the VVO model, more precise measurements influence the decision more strongly, the overall evidence for “no change” becomes stronger and subjects become less likely to report “change”. Our result confirms a prediction from an earlier Bayesian model of change discrimination [28] and provides additional evidence for probabilistic computation by humans in change detection.

### Discussion

We have found that in detecting a change among multiple stimuli: a) the encoding precision of a stimulus is variable even for a given value of stimulus reliability; b) observers near-optimally take into account both variations in stimulus reliability and the additional variability. These results raise several issues.

First, we modeled the distribution of encoding precision as a gamma distribution, with precision being independent across locations and trials. While this choice was convenient and led to good fits, alternatives to the gamma and independence assumptions must be considered.

Second, what causes variability in encoding precision? Several possible factors were mentioned in the introduction. In addition, the precision of memorized items could decay in variable ways, or precision could simply depend on the task-relevant feature value [35]. The relative contributions of these factors remain to be determined.

Third, variability in precision may have implications for encoding models in other tasks. It could potentially account for subject responses that are usually modeled as lapses, since those correspond to a precision of zero. Moreover, in cue combination, it has been suggested that sensory noise is best described by a mixture of a Gaussian and a uniform distribution [36] or of two Gaussian distributions [37]. These mixture models can be regarded as approximations to a full-fledged doubly stochastic process as in Eq. (1), since the mixture components correspond to two different values of precision.

Fourth, how variability in precision can be recognized in neural activity depends on the neural coding scheme one subscribes to. In the framework of Poisson-like probabilistic population codes, variability in encoding precision might correspond to variability in population gain [18,38]. There is initial evidence that gain does vary [12,13,14], and this variability might in part be due to attentional factors [39,40,41,42]. Neuroimaging studies have found that trial-to-trial fluctuations in perceptual performance correlate with fluctuations in stimulus-independent, ongoing neural activity in dorsal anterior cingulate cortex, dorsolateral prefrontal cortex, and dorsal parietal areas [43,44]. This activity might in part reflect the attentional state of the observer, in which case their fluctuations might partially account for variability in precision.

Fifth, how can a neural population “know” encoding precision for use in decision-making? Again in probabilistic population coding, a neural population encodes on each trial a full likelihood function over the stimulus, whose inverse width represents the precision/certainty associated with that stimulus on that trial [18]. Thus, encoding precision is implicitly known on a trial-by-trial basis and can be used in downstream computation. A next step

**Table 1.** List of models considered.

Model	Precision	Local decision variable ( $d_i$ )	Decision rule	#Pars
VVO	variable	$\frac{I_0(\kappa_{x,i})I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(y_i - x_i)}\right)}$	with $\kappa_i$ the actual value at the $i^{\text{th}}$ location $\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^N d_i > 1$	4
VEO	variable	$\frac{I_0(\kappa_{x,i})I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(y_i - x_i)}\right)}$	with $\kappa_i$ either $\kappa_{\text{low}}$ or $\kappa_{\text{high}}$ $\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^N d_i > 1$	4
VAO	variable	$\frac{I_0(\bar{\kappa}_x)I_0(\bar{\kappa}_y)}{I_0\left(\sqrt{\bar{\kappa}_x^2 + \bar{\kappa}_y^2 + 2\bar{\kappa}_x\bar{\kappa}_y\cos(y_i - x_i)}\right)}$	where $\bar{\kappa}$ is an average over locations $\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^N d_i > 1$	4
VSO	variable	$\frac{I_0(\kappa_{\text{assumed}})^2}{I_0(\kappa_{\text{assumed}}\sqrt{2 + 2\cos(y_i - x_i)})}$	$\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^N d_i > 1$	5
VVM	variable	$\frac{I_0(\kappa_{x,i})I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(y_i - x_i)}\right)}$	with $\kappa_i$ the actual value at $\max_i d_i > k$ the $i^{\text{th}}$ location	4
VEM	variable	$\frac{I_0(\kappa_{x,i})I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(y_i - x_i)}\right)}$	with $\kappa_i$ either $\kappa_{\text{low}}$ or $\kappa_{\text{high}}$ $\max_i d_i > k$	4
VAM	variable	$\frac{I_0(\bar{\kappa}_x)I_0(\bar{\kappa}_y)}{I_0\left(\sqrt{\bar{\kappa}_x^2 + \bar{\kappa}_y^2 + 2\bar{\kappa}_x\bar{\kappa}_y\cos(y_i - x_i)}\right)}$	where $\bar{\kappa}$ is an average over $\max_i d_i > k$ locations	4
VSM	variable	$\frac{I_0(\kappa_{\text{assumed}})^2}{I_0(\kappa_{\text{assumed}}\sqrt{2 + 2\cos(y_i - x_i)})}$	$\max_i d_i > k \Leftrightarrow \max_i  y_i - x_i  > \bar{k}$	4
EEO	equal	$\frac{I_0(\kappa_i)^2}{I_0(\kappa_i\sqrt{2 + 2\cos(y_i - x_i)})}$	with $\kappa_i$ either $\kappa_{\text{low}}$ or $\kappa_{\text{high}}$ $\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^N d_i > 1$	3
EAO	equal	$\frac{I_0(\bar{\kappa})^2}{I_0(\bar{\kappa}\sqrt{2 + 2\cos(y_i - x_i)})}$	where $\bar{\kappa}$ is an average over locations $\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^N d_i > 1$	3
ESO	equal	$\frac{I_0(\kappa_{\text{assumed}})^2}{I_0(\kappa_{\text{assumed}}\sqrt{2 + 2\cos(y_i - x_i)})}$	$\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^N d_i > 1$	4
EEM	equal	$\frac{I_0(\kappa_i)^2}{I_0(\kappa_i\sqrt{2 + 2\cos(y_i - x_i)})}$	with $\kappa_i$ either $\kappa_{\text{low}}$ or $\kappa_{\text{high}}$ $\max_i d_i > k$	3
EAM	equal	$\frac{I_0(\bar{\kappa})^2}{I_0(\bar{\kappa}\sqrt{2 + 2\cos(y_i - x_i)})}$	where $\bar{\kappa}$ is an average over $\max_i d_i > k$ locations	3
ESM	equal	$\frac{I_0(\kappa_{\text{assumed}})^2}{I_0(\kappa_{\text{assumed}}\sqrt{2 + 2\cos(y_i - x_i)})}$	$\max_i d_i > k \Leftrightarrow \max_i  y_i - x_i  > \bar{k}$	3

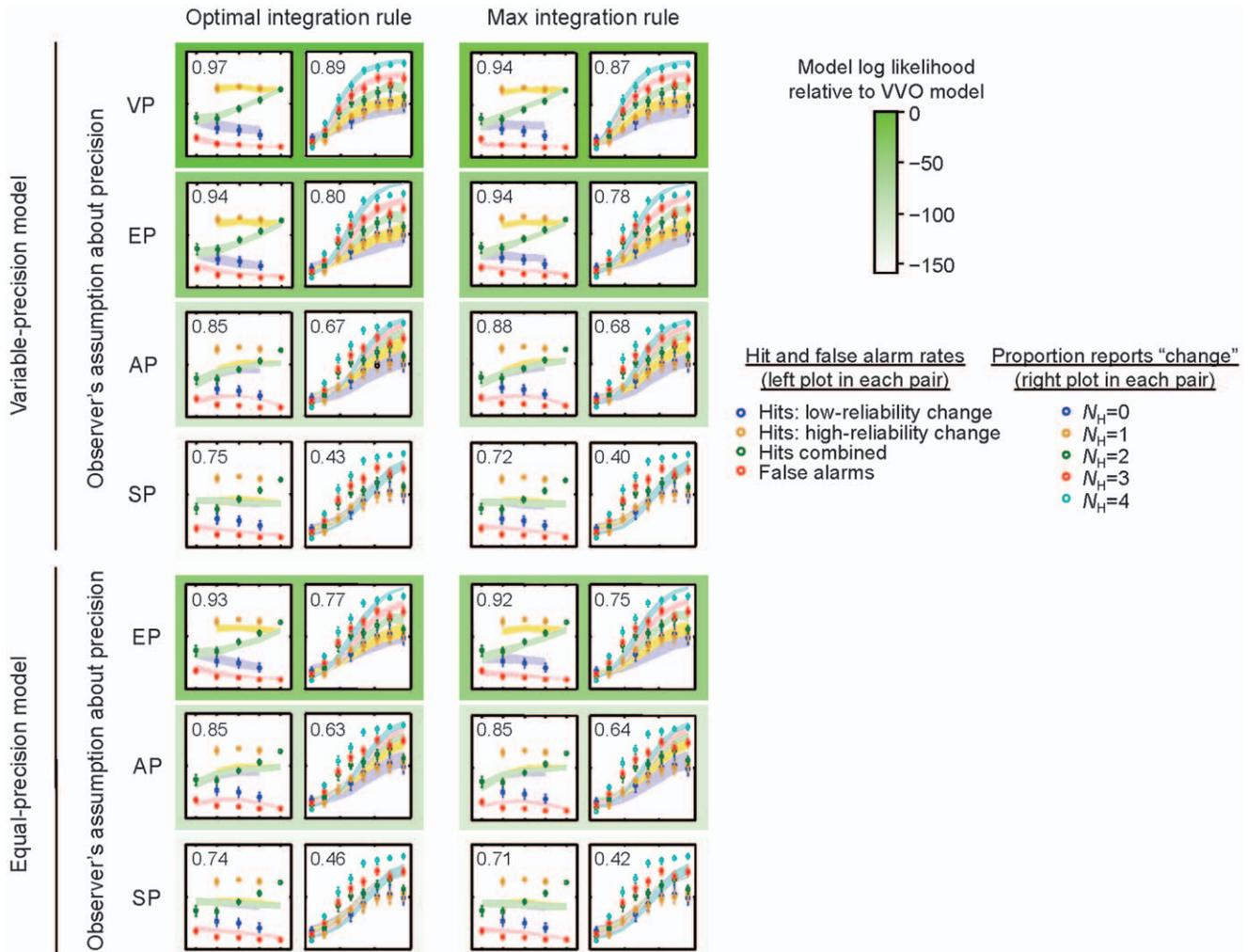
The first letter stands for variable (V) or equal (E) encoding precision. The second letter indicates the observer’s assumption about encoding precision (V: variable; E: equal; A: sample average over locations; S: single value). The third letter stands for the optimal (O) or Max (M) integration rule. The equivalences ( $\Leftrightarrow$ ) in the VSM and FSM models are explained in the Text S1; the notation  $|\cdot|$  denotes circular distance. doi:10.1371/journal.pone.0040216.t001

would be to use probabilistic population codes to design a neural network that takes Poisson-like representations of the individual stimuli in both displays as input and has an output layer that encodes the probability that a change occurred (potentially in the medial temporal lobe [45]).

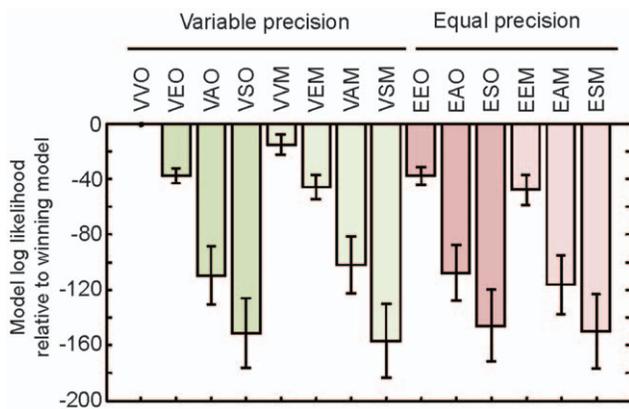
Our work illustrates a new role for change detection in psychology. Traditionally, change detection has only been used to probe capacity limitations in short-term memory [25,46,47,48]. Viewing change detection as inference on noisy sensory measurements is relatively new [33]. Here, we have demonstrated the use of change detection in studying whether the brain computes with probability distributions. Behavioral evidence for probabilistic computation had so far been largely limited to tasks with relatively simple statistical structures, such as cue combination. Change detection is a case study of complex inference, because of the presence of multiple relevant stimuli (spatial complexity), because stimulus information must be integrated into an abstract

categorical judgment (structural complexity), and because perception interacts with visual short-term memory (temporal complexity).

A final caveat. It is tempting to equate optimality with the notion that the brain computes with probabilities on an individual-trial basis (probabilistic computation). These are, however, orthogonal notions [16,49]. In some tasks, such as judging whether an oriented stimulus is tilted to the left or to the right, optimality can be attained using only point estimates and does not require trial-by-trial representations of probability. Conversely, an observer might take into account precision – and perhaps represent probability – on a trial-by-trial and item-by-item basis, but do so in a suboptimal way. Here, we have provided evidence for both optimality and probabilistic computation in change detection. To test for probabilistic computation, we varied reliability unpredictably without giving trial-to-trial feedback, and compared models in which the observer does or does not



**Figure 3. Fits of all 14 models to the data in Fig. 1B-C (axis labels and scales as there).** VP = variable precision; EP = equal precision; AP = average precision; SP = single precision. Error bars and shaded areas represent  $\pm 1$  s.e.m. in the data and the model, respectively. The number in each plot is the  $R^2$  of the fit (for the left plot in each pair, computed over false-alarm rates and unconditioned hit rates). Frame color indicates model goodness of fit relative to the winning model, as obtained from Bayesian model comparison (Fig. 4). doi:10.1371/journal.pone.0040216.g003



**Figure 4. Log likelihood of each model relative to the VVO model.** Negative values indicate that the model is less likely than the VVO model. Error bars represent s.e.m. Abbreviations and color scheme are as in Table 1. doi:10.1371/journal.pone.0040216.g004

take into account precision on a trial-by-trial and item-by-item basis. To test for optimality, we compared the optimal decision rule against a plausible suboptimal one, the Max rule. Thus, we were to some extent able to disentangle Bayesian optimality from probabilistic computation. We speculate that as task complexity increases, optimality will break down at some point, but probabilistic computation continues to be performed – in other words, humans are suboptimal, probabilistic observers.

## Methods

### Stimuli

Stimuli were displayed on a 21" LCD monitor at a viewing distance of 60 cm. Each stimulus display contained four oriented ellipses. Two types of ellipses were used: “long” and “short” ones. “Long” ellipses had minor and major axes of 0.37 and 1.02 degrees of visual angle (deg), respectively. “Short” ellipses had the same area, but their elongations were determined separately for each subject (see Procedure). On each trial, ellipse centers were chosen by placing one at a random location on an imaginary circle of radius 7 deg around the screen center, placing the next one 90°

counterclockwise from the first along the circle, etc., until all four ellipses had been placed. This spacing was sufficiently large to avoid crowding effects. Each ellipse position was jittered by a random amount between  $-0.3$  and  $0.3$  deg in  $x$ - and  $y$ -directions (independently). Stimulus and background luminances were  $95.7$  and  $33.1$  cd/m<sup>2</sup>, respectively.

## Subjects

Nine subjects participated (6 naïve, 3 authors; 1 female). All were between 22 and 32 years old and had normal or corrected-to-normal vision. The study was approved by the Institutional Review Board for Human Subject Research for Baylor College of Medicine; all subjects gave written informed consent.

## Procedure

There were three types of trial blocks: testing blocks, practice blocks, and threshold blocks. In each testing block, a trial began with a blank screen displaying a central fixation cross for 1000 ms. The first stimulus display was presented for 100 ms, followed by a delay period of 1000 ms, followed by a second stimulus display for 100 ms. On each trial, the number of long ellipses was chosen randomly with equal probability from 0 to 4. The locations of the long ellipses were chosen randomly given the constraint of their total number; all other ellipses were short. The orientation of each ellipse was drawn independently from a uniform distribution over all possible orientations. The second stimulus display was identical to the first, except that there was a 50% chance that one of the ellipses had changed its orientation by an angle drawn from a uniform distribution over all possible orientations. Following the second display, the observer pressed a key to indicate whether there was a change between the first and second displays. A response caused the next trial to begin. No trial-by-trial feedback was given. A practice block was identical to a testing block, except that all stimuli on a given trial had the same reliability, which was varied randomly across trials. Stimulus presentation time was initially 333 ms and decreased by 33 ms every 32 trials, allowing the observer to ease into the task. Feedback was given on each trial. The practice session consisted of 256 trials. A threshold block was identical to a practice block but used only the shortest stimulus presentation time (100 ms), and was 400 trials in length.

At the beginning of each session, subjects were informed in lay terms about the distributions from which the stimuli were drawn (e.g., “The change is equally likely to be of any magnitude.”). Each observer completed three sessions on separate days. The first session began with a practice block only for naïve subjects. All subjects then did one threshold block of 400 trials. We fitted a cumulative normal distribution to accuracy as a function of ellipse elongation and extrapolated the performance to the maximal elongation. If the resulting performance was equal to or greater than 75%, we found the elongation of a “short” ellipse from the 65% correct point of the fitted curve. If the resulting extrapolated performance was lower than 75%, the observer repeated the threshold block. If extrapolated performance on the repeated block was again lower than 75%, the observer was excluded from the study. Testing blocks had 400, 800, and 800 testing trials per session, respectively. There were two timed breaks spread evenly for the 400-trial session and four in the 800-trial ones. During each break, a screen showing the percentage correct in the block was displayed. Cumulative performance was shown at the end of each session.

## Encoding Model

For convenience, all orientations were remapped from  $[-\pi/2, \pi/2]$  to  $[-\pi, \pi]$ . For a true stimulus orientation  $\theta$ , we assumed

the measurement  $x$  to follow a Von Mises distribution,  $p(x|\theta; \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\theta)}$ , where  $\kappa$  is the concentration parameter.  $\kappa$  is determined by the amount of resource allocated to the stimulus,  $\mathcal{J}$ . The relationship between  $\mathcal{J}$  and  $\kappa$  is  $J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)}$ , where  $I_1$  is the modified Bessel function of the first kind of order 1 (see Text S1). In the EP model,  $\mathcal{J}$  is determined by ellipse elongation only. In the VP model,  $\mathcal{J}$  is drawn from a gamma distribution with mean  $\bar{J}$  and scale parameter  $\tau$ , where  $\bar{J}$  is determined by ellipse elongation (it is accordingly denoted  $\bar{J}_{\text{low}}$  or  $\bar{J}_{\text{high}}$ ).

## Model Predictions

We are interested in computing the probability predicted by a model of reporting “change” for a set of stimuli and corresponding reliabilities, given a set of parameter values. This probability is equal to the probability that  $d > 1$  for measurements  $(\mathbf{x}, \mathbf{y})$  drawn using the generative model with the given parameters. This probability only depends on the magnitude of change,  $\Delta$ , the number of high-reliability stimuli,  $N_H$ , and whether a change, if any, occurred in a low-reliability or a high-reliability stimulus. We binned  $\Delta$  every 3 degrees between 0 and 90 degrees, resulting in 31 values;  $N_H$  takes 5 possible values, resulting in  $31 \cdot 5 \cdot 2 = 310$  trial types. For each trial type, we approximated the distributions of  $\mathbf{x}$  and  $\mathbf{y}$  using a Monte Carlo simulation with 1,000 samples. For each sample, the model’s decision rule was applied, and the proportion of “change” responses among all samples was determined. This returned an estimate of the model’s probability of reporting “change” on a given trial, for the given parameter values. The entire procedure was repeated for all parameter combinations.

## Model Fitting

For a given model, we denote the vector of model parameters by  $\mathbf{t}$ . The likelihood of  $\mathbf{t}$  is the probability of the human subject’s empirical responses given  $\mathbf{t}$ :

$$L(\mathbf{t}) = p(\text{data}|\mathbf{t}, \boldsymbol{\theta}) = \prod_{k=1}^{N_{\text{trials}}} p(\hat{C}_k | \text{stimuli}_k, \mathbf{t}),$$

where  $N_{\text{trials}}$  is the total number of trials,  $\hat{C}_k$  the subject’s response on the  $k^{\text{th}}$  trial, and  $\text{stimuli}_k$  is shorthand for the stimulus orientations and their reliabilities in both displays. The maximum-likelihood estimate of the parameters is the value of  $\mathbf{t}$  that maximizes  $L(\mathbf{t})$ .

## Bayesian Model Comparison

Each model  $m$  produces a prediction about the response on each trial,  $p(\hat{C}_k | \text{stimuli}_k, \mathbf{t}, m)$ . Bayesian model comparison [34] consists of calculating for each model the probability of finding a subject’s actual responses under this distribution, averaged over free parameters:

$$L(m) = p(\text{data}|m) = \int p(\text{data}|m, \mathbf{t}) p(\mathbf{t}|m) d\mathbf{t} \\ = \int \left( \prod_{k=1}^{N_{\text{trials}}} p(\hat{C}_k | \text{stimuli}_k, \mathbf{t}, m) \right) p(\mathbf{t}|m) d\mathbf{t}$$

It is convenient to compute the logarithm of  $L(m)$  and write it as.

$$\log L(m) = L^*(m) + \log \int e^{L(t,m) - L^*(m)} p(\mathbf{t}|m) d\mathbf{t}, \quad (4)$$

where  $L(\mathbf{t}, m) = \sum_{k=1}^{N_{\text{trials}}} \log p(\hat{C}_k | \text{stimuli}_k, \mathbf{t}, m)$  and  $L^*(m) = \max_{\mathbf{t}} L(\mathbf{t}, m)$ . This form prevents numerical problems, since the exponential in the integrand of Eq. (4) is now of order 1 near the maximum-likelihood value of  $\mathbf{t}$ . For the parameter prior, we assume a uniform distribution across some range, whose size we denote  $R_j$  for the  $j^{\text{th}}$  parameter. Ranges were as follows:  $[1, 100]$  for  $\bar{J}_{\text{low}}$ ,  $\bar{J}_{\text{high}}$ ,  $\bar{J}_{\text{assumed}}$ ,  $\bar{J}_{\text{low}}$ , and  $\bar{J}_{\text{high}}$ ;  $[1, 30]$  for  $\tau$ ;  $[-2.2, 51.8]$  for the Max model criterion  $k$ ;  $[0.3, 0.7]$  for  $p_{\text{change}}$ . Eq. (4) becomes  $\log L(m) = L^*(m) - \sum_{j=1}^{\dim \mathbf{t}} \log R_j + \log \int e^{L(\mathbf{t}, m) - L^*(m)} d\mathbf{t}$ , where  $\dim \mathbf{t}$  is the number of parameters. We approximated the integral through a Riemann sum. We tested the parameter fitting and model comparison code on fake data generated from each of the 14 models; parameters were estimated correctly and the model used to generate the data always won, showing that the models are distinguishable using this method.

## References

- Tolhurst D, Movshon J, Dean A (1982) The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research* 23: 775–785.
- Faisal A, Selen LPJ, Wolpert DM (2008) Noise in the nervous system. *Nat Rev Neurosci* 9: 292–303.
- Gershon ED, Wiener MC, Latham PE, Richmond BJ (1998) Coding strategies in monkey V1 and inferior temporal cortices. *Journal of Neurophysiology* 79: 1135–1144.
- Shadlen MN, Britten KH, Newsome WT, Movshon JA (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* 16: 1486–1510.
- Green DM, Swets JA (1966) Signal detection theory and psychophysics. Los Altos, CA: John Wiley & Sons.
- Matthias E, Bublak P, Costa A, Mueller HJ, Schneider WX, et al. (2009) Attentional and sensory effects of lowered levels of intrinsic alertness. *Neuropsychologia* 47: 3255–3264.
- Brady TF, Tenenbaum JB (2010) Encoding higher-order structure in visual working memory: A probabilistic model. In: Ohlsson S, Catrambone R, editors. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science. 411–416.
- Brady TJ, Alvarez GA (2011) Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psych Science* 22: 384–392.
- Posner MI (1980) Orienting of attention. *Q J Exp Psychol* 32: 3–25.
- Pestilli F, Carrasco M (2005) Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research* 45: 1867–1875.
- Van den Berg R, Shin H, Chou W-C, George R, Ma WJ (2012) Variability in encoding precision accounts for visual short-term memory limitations. *Proc Natl Acad Sci U S A* 109: published online May 11.
- Goris RLT, Simoncelli EP, Movshon JA (2012) Using a doubly-stochastic model to analyze neuronal activity in the visual cortex. *Cosyne Abstracts*. Salt Lake City.
- Churchland AK, Kiani R, Chaudhuri R, Wang X-J, Pouget A, et al. (2011) Variance as a signature of neural computations during decision-making. *Neuron* 69: 818–831.
- Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci* 13: 369–378.
- Cohen MR, Maunsell JHR (2010) A neuronal population measure of attention predicts behavioral performance on individual trials. *J Neurosci* 30: 15241–15253.
- Ma WJ (2010) Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Research* 50: 2308–2319.
- Pouget A, Dayan P, Zemel RS (2003) Inference and computation with population codes. *Annual Review of Neuroscience* 26: 381–410.
- Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9: 1432–1438.
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429–433.
- Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27: 712–719.
- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14: 257–262.
- Ma WJ, Navalpakkam V, Beck JM, Van den Berg R, Pouget A (2011) Behavior and neural basis of near-optimal visual search. *Nat Neurosci* 14: 783–790.
- Van den Berg R, Vogel M, Josic K, Ma WJ (2011) Optimal inference of sameness. *Proc Natl Acad Sci U S A* 109: 3178–3183.
- French RS (1953) The discrimination of dot patterns as a function of number and average separation of dots. *J Exp Psychol* 46: 1–9.
- Pashler H (1988) Familiarity and visual change detection. *Percept Psychophys* 44: 369–378.
- Phillips WA (1974) On the distinction between sensory storage and short-term visual memory. *Percept Psychophys* 16: 283–290.
- Yule GU (1903) Notes on the theory of association of attributes in statistics. *Biometrika* 2: 121–134.
- Ma WJ, Huang W (2009) No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *J Vision* 9: 3: 1–30.
- Nolte LW, Jaarsma D (1966) More on the detection of one of  $M$  orthogonal signals. *J Acoust Soc Am* 41: 497–505.
- Palmer J, Verghese P, Pavel M (2000) The psychophysics of visual search. *Vision Research* 40: 1227–1268.
- Eckstein MP, Thomas JP, Palmer J, Shimozaki SS (2000) A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Percept Psychophys* 62: 425–451.
- Baldassi S, Verghese P (2002) Comparing integration rules in visual search. *J Vision* 2: 559–570.
- Wilken P, Ma WJ (2004) A detection theory account of change detection. *J Vision* 4: 1120–1135.
- MacKay DJ (2003) *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.
- Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* 14: 926–932.
- Natarajan R, Murray I, Shams L, Zemel RS (2009) Characterizing response behavior in multisensory perception with conflicting cues. *Adv Neural Information Processing Systems* 21. Cambridge, MA: MIT Press.
- Girshick AR, Banks MS (2009) Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts. *J Vision* 9: 8:1–20.
- Seung H, Sompolinsky H (1993) Simple model for reading neuronal population codes. *Proceedings of National Academy of Sciences USA* 90: 10749–10753.
- Saproo S, Serences JT (2010) Spatial attention improves the quality of population codes in human visual cortex. *J Neurophys* 104: 885–895.
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18: 193–222.
- Connor CE, Gallant JL, Preddie DC, Van Essen DC (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. *J Neurophysiol* 75: 1306–1308.
- McAdams CJ, Maunsell JH (1999) Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron* 23: 765–773.

## Supporting Information

**Figure S1 Encoding precision as a function of the concentration parameter of the Von Mises distribution.**

The dashed line is the identity line.

(TIFF)

**Table S1 Parameter values for all models.** Mean and s.e.m. are over subjects.

(DOCX)

**Text S1 Supporting information.** Contains: Relationship between precision and concentration parameter; Equal-precision and variable-precision models; Optimally inferring change occurrence; The Max model

(DOCX)

## Author Contributions

Conceived and designed the experiments: SK RvdB WJM. Performed the experiments: SK. Analyzed the data: SK RvdB. Wrote the paper: SK RvdB WJM. Developed the theory: SK RvdB WJM.

43. Sadaghiani S, Hesselmann G, Friston KJ, Kleinschmidt A (2010) The relation of ongoing brain activity, evoked neural responses, and cognition. *Front Syst Neurosci* 4: 20.
44. Sapiro A, d'Avossa G, McAvoy M, Shulman GL, Corbetta M (2005) Brain signals for spatial attention predict performance in a motion discrimination task. *Proc Natl Acad Sci U S A* 102: 17810–17815.
45. Reddy L, Quiñan Quiroga R, Wilken P, Koch C, Fried I (2006) A single-neuron correlate of change detection and change blindness in the human medial temporal lobe. *Curr Biol* 2006: 20.
46. Eng HY, Chen D, Jiang Y (2005) Visual working memory for simple and complex visual stimuli. *Psychon B Rev* 12: 1127–1133.
47. Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24: 87–114.
48. Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390: 279–281.
49. Whiteley L, Sahani M (2008) Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *J Vision* 8: 1–15.

# Probabilistic computation in human perception under variability in encoding precision

## Supporting Information

Shaiyan Keshvari, Ronald van den Berg, Wei Ji Ma  
Baylor College of Medicine, Houston, TX, USA

### 1 Relationship between precision and concentration parameter

For convenience, we remap all orientations from  $[-\pi/2, \pi/2)$  to  $[-\pi, \pi)$ . We assume that the measurement  $x$  of a stimulus orientation  $\theta$  follows a Von Mises distribution with concentration parameter  $\kappa$ :

$$p(x|\theta;\kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\theta)}. \quad (\text{S.1})$$

We identify encoding precision with Fisher information,  $J(\theta)$ , which measures the performance of the best possible unbiased decoder. Substituting Eq. (S.1) into the definition of Fisher information, we find

$$J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)}. \quad (\text{S.2})$$

$J$  is a monotonically increasing function of  $\kappa$  and therefore invertible. It is shown in Fig. S1. The equivalent relationship for a Gaussian distribution would be  $J=1/\sigma^2$ .

### 2 Equal-precision and variable-precision models

In the equal-precision (EP) model, the encoding precision of an observation,  $J$ , is determined only by the elongation of the ellipse producing the observation. In the variable-precision (VP) model, encoding precision is a random variable that follows, independently for each item and each trial, a gamma distribution with mean  $\bar{J}$  and scale parameter  $\tau$ , denoted  $p(J|\bar{J};\tau)$ .

Variance is equal to  $\bar{J}\tau$ . The mean  $\bar{J}$  differs between a long and a short ellipse. Thus, in this model, a measurement  $x$  is drawn from a doubly stochastic process: first precision is drawn from  $p(J|\bar{J};\tau)$ , then the measurement from the Von Mises distribution in Eq. (S.1) with  $\kappa$  determined by  $J$  through Eq. (S.2).

### 3 Optimally inferring change occurrence

The generative model (statistical structure) of the task is shown in Fig. 2c. We denote by  $C$  the binary variable of change occurrence (0: no change; 1: change), with a true prior  $p(C)=0.5$ . The orientations in the first display,  $\theta=(\theta_1,\dots,\theta_N)$ , are drawn independently from a uniform distribution over orientation space, and therefore we have

$$p(\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^N.$$

We denote by  $\Delta$  the vector of change magnitudes at all locations. It is a vector of zeros except at the location where the change, of magnitude  $\Delta$ , is located. Since all locations are equally likely to contain the change, we have

$$p(\Delta | C, \Delta) = \frac{1}{N} \sum_{i=1}^N \delta(\Delta - C\Delta \mathbf{1}_i),$$

where  $\mathbf{1}_i$  is the vector that has a 1 in the  $i^{\text{th}}$  entry and 0 everywhere else. The magnitude of change is drawn from a uniform distribution,  $p(\Delta) = 1/(2\pi)$ . The orientations in the second display,  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_N)$ , are the sum of the initial orientations  $\boldsymbol{\theta}$  and the changes  $\Delta$ :

$$p(\boldsymbol{\varphi} | \Delta, \boldsymbol{\theta}) = \delta(\boldsymbol{\varphi} - (\Delta + \boldsymbol{\theta})).$$

Finally, the measurements in the first and second displays,  $\mathbf{x} = (x_1, \dots, x_N)$  and  $\mathbf{y} = (y_1, \dots, y_N)$ , respectively, are drawn from independent Von Mises distributions with corresponding concentration parameters  $\boldsymbol{\kappa}_x = (\kappa_{x,1}, \dots, \kappa_{x,N})$  and  $\boldsymbol{\kappa}_y = (\kappa_{y,1}, \dots, \kappa_{y,N})$ :

$$p(\mathbf{x} | \boldsymbol{\theta}; \boldsymbol{\kappa}_x) = \prod_{i=1}^N p(x_i | \theta_i; \kappa_{x,i}) = \prod_{i=1}^N \frac{1}{2\pi I_0(\kappa_{x,i})} e^{\kappa_{x,i} \cos(x_i - \theta_i)}$$

$$p(\mathbf{y} | \boldsymbol{\varphi}; \boldsymbol{\kappa}_y) = \prod_{i=1}^N p(y_i | \varphi_i; \kappa_{y,i}) = \prod_{i=1}^N \frac{1}{2\pi I_0(\kappa_{y,i})} e^{\kappa_{y,i} \cos(y_i - \varphi_i)}$$

We have to consider the possibility that  $\kappa_{x,i} \neq \kappa_{y,i}$ , because in the VP model, both sets of concentration parameters are independent random variables.

We are now ready to specify the observer's inference process. Optimal inference amounts to "inverting" the generative model, i.e., to compute the probability of  $C$  given the noisy measurements of the orientations in the first and second displays on a given trial,  $\mathbf{x}$  and  $\mathbf{y}$ . The observer's decision variable  $d$  is defined as the ratio between the probability that a change occurred and the probability that no change occurred, given  $\mathbf{x}$  and  $\mathbf{y}$ :

$$d = \frac{p(C=1 | \mathbf{x}, \mathbf{y})}{p(C=0 | \mathbf{x}, \mathbf{y})} = \frac{p(\mathbf{x}, \mathbf{y} | C=1) p(C=1)}{p(\mathbf{x}, \mathbf{y} | C=0) p(C=0)}, \quad (\text{S.3})$$

where  $p(C)$  is the prior over  $C$ . The ideal observer will respond "change" ( $\hat{C}=1$ ) if  $d > 1$  and "no change" ( $\hat{C}=0$ ) when  $d < 1$ . The joint distribution of  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $C$  is obtained by marginalizing the

joint distribution of all variables over  $\Delta$ ,  $\mathbf{\Delta}$ ,  $\boldsymbol{\theta}$ , and  $\boldsymbol{\varphi}$ . It is most convenient to first marginalize over  $\Delta$ ,  $\mathbf{\Delta}$ , and substitute their distributions:

$$\begin{aligned}
p(\mathbf{x}, \mathbf{y}, C) &= \iiint p(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \Delta, \mathbf{\Delta}, C) d\Delta d\mathbf{\Delta} d\boldsymbol{\theta} d\boldsymbol{\varphi} \\
&= \iint p(\mathbf{x}, \mathbf{y} | \Delta) p(\mathbf{\Delta} | C, \Delta) p(\Delta) p(C) d\Delta d\mathbf{\Delta} \\
&= \iint p(\mathbf{x}, \mathbf{y} | \Delta) \frac{1}{N} \sum_{i=1}^N \delta(\Delta - C\mathbf{\Delta}\mathbf{1}_i) \frac{1}{2\pi} p(C) d\Delta d\mathbf{\Delta} \\
&= \frac{1}{2\pi} \frac{p(C)}{N} \sum_{i=1}^N \iint p(\mathbf{x}, \mathbf{y} | \Delta = C\mathbf{\Delta}\mathbf{1}_i) d\Delta \\
&= \frac{1}{2\pi} \frac{p(C)}{N} \sum_{i=1}^N \iint p(x_i, y_i | \Delta_i = C\Delta) \left( \prod_{j \neq i} p(x_j, y_j | \Delta_j = 0) \right) d\Delta,
\end{aligned} \tag{S.4}$$

where in the last equality we have used the fact that given the vector  $\mathbf{\Delta}$ , the pairs of observations  $(x_i, y_i)$  are conditionally independent across locations. Substituting Eq. (S.4) into Eq. (S.3) and denoting the prior probability of  $C=1$  by  $p_{\text{change}}$  gives

$$\begin{aligned}
d &= \frac{p_{\text{change}}}{1 - p_{\text{change}}} \frac{\frac{1}{2\pi} \frac{p(C)}{N} \sum_{i=1}^N \int p(x_i, y_i | \Delta_i = \Delta) \left( \prod_{j \neq i} p(x_j, y_j | \Delta_j = 0) \right) d\Delta}{\frac{1}{2\pi} \frac{p(C)}{N} \sum_{i=1}^N \int \left( \prod_{i=1}^N p(x_i, y_i | \Delta_i = 0) \right) d\Delta} \\
&= \frac{p_{\text{change}}}{1 - p_{\text{change}}} \frac{\frac{1}{2\pi N} \sum_{i=1}^N \int p(x_i, y_i | \Delta_i = \Delta) \left( \prod_{j \neq i} p(x_j, y_j | \Delta_j = 0) \right) d\Delta}{\prod_{i=1}^N p(x_i, y_i | \Delta_i = 0)}.
\end{aligned}$$

We now notice that the products in the numerator and denominator almost cancel, leaving only the  $i^{\text{th}}$  term:

$$d = \frac{p_{\text{change}}}{1 - p_{\text{change}}} \frac{1}{N} \sum_{i=1}^N \frac{\int p(x_i, y_i | \Delta_i = \Delta) d\Delta}{2\pi p(x_i, y_i | \Delta_i = 0)}.$$

The summand is exactly equal to the log likelihood ratio of a change occurring at the  $i^{\text{th}}$  location, so we can write:

$$\boxed{d = \frac{p_{\text{change}}}{1 - p_{\text{change}}} \frac{1}{N} \sum_{i=1}^N d_i}, \tag{S.5}$$

where the local decision variable  $d_i$  is defined as

$$d_i = \frac{\int p(x_i, y_i | \Delta_i = \Delta) d\Delta}{2\pi p(x_i, y_i | \Delta_i = 0)}.$$

We can evaluate  $d_i$  by marginalizing over  $\theta_i$  and  $\varphi_i$  and substituting the Von Mises noise distributions:

$$\begin{aligned} d_i &= \frac{\int p(x_i, y_i | \Delta_i = \Delta) d\Delta}{2\pi p(x_i, y_i | \Delta_i = 0)} \\ &= \frac{\iiint p(x_i, y_i | \theta_i, \varphi_i) p(\theta_i) p(\varphi_i | \theta_i, \Delta_i = \Delta) d\theta_i d\varphi_i d\Delta}{2\pi \iint p(x_i, y_i | \theta_i, \varphi_i) p(\theta_i) p(\varphi_i | \theta_i, \Delta_i = 0) d\theta_i d\varphi_i} \\ &= \frac{\iint p(x_i | \theta_i) p(y_i | \theta_i + \Delta) d\theta_i d\Delta}{2\pi \int p(x_i | \theta_i) p(y_i | \theta_i) d\theta_i} \\ &= \frac{\iint \frac{1}{2\pi I_0(\kappa_{x,i})} e^{\kappa_{x,i} \cos(x_i - \theta_i)} \frac{1}{2\pi I_0(\kappa_{y,i})} e^{\kappa_{y,i} \cos(y_i - \theta_i - \Delta)} d\theta_i d\Delta}{2\pi \int \frac{1}{2\pi I_0(\kappa_{x,i})} e^{\kappa_{x,i} \cos(x_i - \theta_i)} \frac{1}{2\pi I_0(\kappa_{y,i})} e^{\kappa_{y,i} \cos(y_i - \theta_i)} d\theta_i} \\ &= \frac{1}{2\pi \int \frac{1}{2\pi I_0(\kappa_{x,i})} e^{\kappa_{x,i} \cos(x_i - \theta_i)} \frac{1}{2\pi I_0(\kappa_{y,i})} e^{\kappa_{y,i} \cos(y_i - \theta_i)} d\theta_i} \\ &= \frac{I_0(\kappa_{x,i}) I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2} + 2\kappa_{x,i} \kappa_{y,i} \cos(y_i - x_i)\right)}. \end{aligned} \tag{S.6}$$

Substituting this result back into Eq. (S.5), we arrive at our final expression for the optimal decision variable:

$$d = \frac{1}{N} \frac{p_{\text{change}}}{1 - p_{\text{change}}} \sum_{i=1}^N \frac{I_0(\kappa_{x,i}) I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2} + 2\kappa_{x,i} \kappa_{y,i} \cos(y_i - x_i)\right)}. \tag{S.7}$$

The observer reports “change” when  $d$  exceeds 1. As one would expect, the decision depends on the circular differences between the measurements in the first and second displays,  $y_i - x_i$ . The local concentration parameters in both displays,  $\kappa_{x,i}$  and  $\kappa_{y,i}$ , weight – albeit in a way that looks a bit less familiar than in cue combination with Gaussian distributions – the contribution of the evidence at each location to the decision.

The suboptimal assumptions about precision – average precision and single precision, as well as equal precision when precision is in reality variable – enter through modifications of Eq. (S.7). For the average-precision assumption, each  $\kappa_{x,i}$  is replaced by an average over locations,  $\bar{\kappa}_x$ , and similarly each  $\kappa_{y,i}$  by  $\bar{\kappa}_y$ . These numbers still vary from trial by trial and might be interpreted as a “gist” representation of precision. (We average concentration parameters  $\kappa$  rather than precision values  $J$  for convenience, but in view of the nearly linear relationship shown in Fig. S1, these are almost equivalent.) For the single-precision assumption, each  $\kappa_{x,i}$  and  $\kappa_{y,i}$  in the decision variable, Eq. (S.7), are replaced by a single assumed concentration parameter,  $\kappa_{\text{assumed}}$ .

#### 4 The Max model

The decision rule in the Max model is  $\max_i d_i > k$ , where  $d_i$  is the local decision variable from Eq. (S.6) and  $k$  is a criterion. When the local decision variable is modified using the assumption that the encoding precision of a stimulus has a single value throughout the experiment (which is a wrong assumption in our experiment already because stimulus reliability can be low or high), then the Max decision rule simplifies. This is because  $\max_i d_i$  can be rewritten as

$$\max_i \frac{I_0(\kappa_{\text{assumed}})^2}{I_0(\kappa_{\text{assumed}} \sqrt{2 + 2 \cos(y_i - x_i)})} = \frac{I_0(\kappa_{\text{assumed}})^2}{I_0(\kappa_{\text{assumed}} \sqrt{2 + 2 \min_i \cos(y_i - x_i)})},$$

where we have used the facts that  $\kappa_{\text{assumed}}$  is a constant and  $I_0$  is a monotonically increasing function. The resulting rule can be rewritten as a condition on  $\min_i \cos(y_i - x_i)$ :

$$\min_i \cos(y_i - x_i) < \tilde{k} \tag{S.8}$$

where  $\tilde{k} = \frac{1}{2} \left( \frac{1}{\kappa_{\text{assumed}}} I_0^{-1} \left( \frac{I_0(\kappa_{\text{assumed}})^2}{k} \right) \right)^2 - 1$  is a different criterion. This proves the

equivalences in Table 1. Note that  $\kappa_{\text{assumed}}$  and  $k$  both only appear in the model through this decision rule, and can therefore not be fitted independently: they are effectively a single parameter. Eq. (S.8) is equivalent to comparing the largest circular distance between two observations at the same location to a criterion:

$$\max_i |y_i - x_i| > \tilde{k},$$

where  $|\cdot|$  denotes the unsigned distance along the circle. In earlier work, we called this suboptimal model the *maximum-absolute-differences model*<sup>1</sup>. In that paper, stimulus reliability

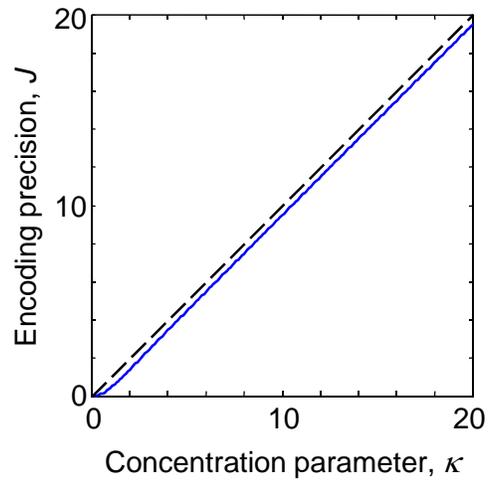
was not deliberately varied and this model would be closer to optimal than it is in the present experiment.

## References

1. Wilken, P. & Ma, W.J. A detection theory account of change detection. *J Vision* **4**, 1120-1135 (2004).

**Table S1: Parameter values for all models.** Mean and s.e.m. are over subjects.

	Variable-precision		Fixed-precision	
	Optimal	Max	Optimal	Max
<b>VP assumption</b>				
$J_{low}$	$6.6 \pm 1.4$	$7.5 \pm 0.8$		
$J_{high}$	$22 \pm 3$	$21 \pm 1$		
$\tau$	$21 \pm 2$	$24 \pm 3$		
$k$	N/A	$-0.44 \pm 0.06$		
$p_{change}$	$0.47 \pm 0.02$	N/A		
<b>FP assumption</b>				
$J_{low}$	$1.6 \pm 0.2$	$2.8 \pm 0.7$	$1.3 \pm 0.1$	$1.5 \pm 0.1$
$J_{high}$	$7.9 \pm 1.6$	$9.8 \pm 3.6$	$5.3 \pm 1.2$	$5.4 \pm 1.2$
$\tau$	$1.7 \pm 0.7$	$2.7 \pm 1.2$	N/A	N/A
$k$	N/A	$0.94 \pm 0.44$	N/A	$-0.11 \pm 0.07$
$p_{change}$	$0.36 \pm 0.02$	N/A	$0.43 \pm 0.01$	N/A
<b>AP assumption</b>				
$J_{low}$	$2.5 \pm 0.3$	$9.5 \pm 2.0$	$1.8 \pm 0.1$	$1.9 \pm 0.2$
$J_{high}$	$5.0 \pm 0.5$	$18 \pm 5$	$3.5 \pm 0.3$	$3.5 \pm 0.3$
$\tau$	$1.3 \pm 0.2$	$16 \pm 4$	N/A	N/A
$k$	N/A	$5.1 \pm 1.2$	N/A	$0.01 \pm 0.09$
$p_{change}$	$0.32 \pm 0.01$	N/A	$0.39 \pm 0.01$	N/A
<b>SP assumption</b>				
$J_{low}$	$3.5 \pm 0.3$	$3.3 \pm 0.2$	$2.1 \pm 0.2$	$2.0 \pm 0.1$
$J_{high}$	$4.2 \pm 0.4$	$4.3 \pm 0.4$	$2.7 \pm 0.4$	$2.7 \pm 0.2$
$\tau$	$1.5 \pm 0.4$	$1.6 \pm 0.2$	N/A	N/A
$J_{assumed}$	$1.5 \pm 0.4$	N/A	$1.9 \pm 0.2$	N/A
$k$	N/A	$-0.86 \pm 0.08$	N/A	$-0.86 \pm 0.08$
$p_{change}$	$0.50 \pm 0.01$	N/A	$0.44 \pm 0.03$	N/A



**Fig. S1**, Encoding precision  $J$  as a function of the concentration parameter of the Von Mises distribution. The dashed line is the identity line.